

データ品質 / ガバナンス

AI時代に求められる データ価値の最大化

ガバナンスを強化して
ビジネス成果を加速するデータ製品

今日の急速に変化するビジネス環境において、データは重要な戦略的資産です。また、業務の最適化・イノベーションの推進・成長の加速を支える基盤となります。データはただ数値や統計を示すだけでなく、成功への道筋を示すインサイト・傾向・パターンを明らかにします。生成 AI と機械学習がこれまで以上に注目を集める中、企業は根本的な課題に取り組んでいます。

データの信頼性とガバナンスを強化し、迅速かつ最大限にデータの価値を引き出してビジネス成果を加速するには、どうすればよいのか？

この課題を解決する鍵は、スピードです。最近の IDC 社の調査では、意思決定者の 75% が、データの価値は数日以内に失われると回答しています。¹

膨大なデータを管理するだけでは不十分です。情報を業務領域に適した高品質な資産へと変換し、即座にアクセスできるようにすることが求められます。データを扱う人なら、実現が非常に困難であることは理解できるでしょう。データと分析のリーダーは、包括的なデータ戦略の推進を妨げるいくつかの課題に直面しています。

- **プロジェクト指向のデータ管理における拡張性の欠如:**各プロジェクトチームは、プロジェクトの範囲・目的・関係者など、独自のパラメーターセットの範囲内で業務を遂行します。このように分断された方法は、部門間のコラボレーションを妨げ、データからの価値の創出を制限します。また、プロジェクト間の関係者の対立、作業の重複、ガバナンスの相違、非効率的なデータ利用、コストの増大、ビジネス成果の遅延や消失といった多岐にわたるリスクがあります。

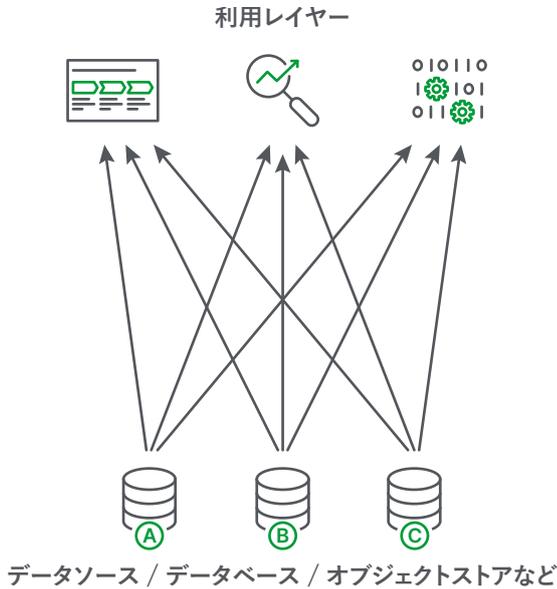
- 全社にわたるシステム間のデータ品質の一貫性の欠如:意思決定プロセスを妨げ、データに基づいたインサイトの信頼性が低下します。
- データと価値のサプライチェーン全体にわたるデータの作成者と利用者の明確な所有権と説明責任の欠如:ビジネスのスピードでデータから価値を創出することが困難になります。
- 分断されたソリューションと独自の一時的なデータ管理法への依存:価値実現に要する時間が長くなります。また、ベテランの人材不足で、拡張や保守に要するコストが増大します。

こうした課題は、データの作成者と利用者との間の溝が深まる原因になります。IT を重視しているデータ作成者は、データの使用方法やデータ利用者のニーズへの理解は二の次で、データプラットフォームを優先します。一方、データ利用者は、自身の業務領域を重視しているため、データの要件をわかりやすく提示することに苦慮しています。さらに、提供されたデータを信頼していません。双方の間の深まる溝を解消するには、戦略的かつ効率的にデータを管理する必要があります。

¹ IDC Blog「Navigating the Planes of Enterprise Intelligence Architecture」2023年 6月

データ管理法の転換

企業は、データから価値を引き出すために、**分断された方法**と**一元管理された方法**という、2つの異なる方法でデータを管理してきました。



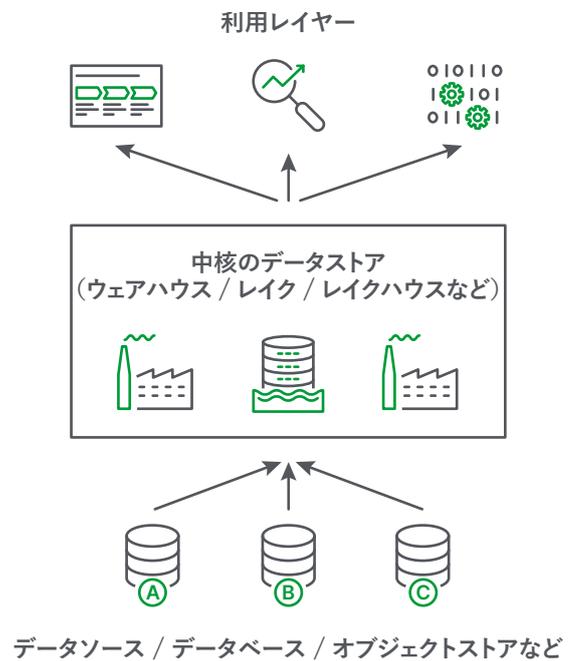
分断された方法

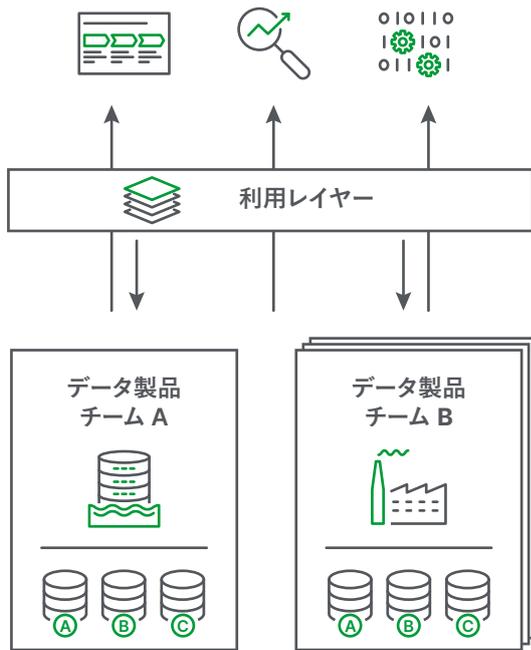
個々の「データプロジェクト」チームが、あらゆるソースシステムからデータを収集し、特定のビジネスユースケースに対処します。各チームはデータを単独で使用し、独自のツールとテクノロジーで、特定のニーズに沿ってデータを抽出・統合・分析します。短期的には即座に価値を得ることができますが、長期的にはビジネスの成長に合わせて拡張することができません。そのため、データのサイロ化が進み、作業の重複やコストの増大で、ビジネスの成長を妨げます。

一元管理された方法

単一の中央データチームが、大規模なデータの抽出・クレンジング・集約を担当します。中央データチームの役割は、組織の分析要件に包括的に対応することです。複雑なデータパイプラインを構築し、あらゆるシステムからデータを取得してクレンジングし、クラウドベースのデータレイク・データウェアハウス・データレイクハウス (Databricks / Delta Lake / Snowflake / Big Query / S3 など) 内で、データを統合します。

この方法は、分断された方法を解決できるように思えるかもしれませんが、致命的な欠点があります。変更や更新には、部門を超えた広範な調整が必要になるため、組織の俊敏性が低下します。また、中央データチームが全社共通のデータ要件を重視すると、ビジネス関係者やデータ利用者の特定のニーズや嗜好を見落としてしまいがちです。そのため、ビジネス要件に合致しないソリューションを提案する可能性があります。





連携された方法

従来の分断された方法や一元管理された方法には限界があるため、企業のデータ利用のニーズに応えることはできません。こうした課題に対応できるようデータ管理の方法に大きな変化が起きています。急速に普及している**連携された方法**は、プラットフォームのエンジニアリング機能を一元管理しながらデータ製品を活用します。これにより、データの提供を戦略的に分散することができます。

複数のデータ製品チームが並行して作業を行い、業務領域に特化したデータ製品に集中することができます。この方法は、ビジネス部門と技術部門を横断したコラボレーションを可能にし、ビジネス成果を加速することができます。

以下は、連携された方法の主なメリットです。



俊敏性と柔軟性

エンドツーエンドのデータ管理を効率化し、データから**迅速にビジネス価値を引き出す**ことができます。製品主導型の考え方を採用し、データの提供と価値の創出を部門を横断した製品チームに分散して俊敏性を高めます。一方、プラットフォームエンジニアリング機能は集約して実行します。これにより、プロセス全体でデータの整合性と説明責任を確保しながら、**変化するビジネスニーズに迅速に対応**できるようになります。



優れたコスト効率

連携されたモデルは、**再利用**が可能のため、投資を最大限に活用して運用コストを削減することができます。さらに、異なる製品チーム間で基本的な構成を集約して共有することで、**余計な作業を最小限に抑えて長期的な保守性を高める**ことができます。



ユースケースの提供を加速

部門を横断したチーム間のコラボレーションを促進し、変化するビジネス要件やデータ主導型の新たな取り組みに適応します。これにより、**特定の業務領域で迅速に成果を挙げ**ることができます。複数のデータ製品チームを横断して並行して開発を実行することで、**拡張性を確保**します。これにより、**業務領域に応じたユースケースのカスタマイズ**が可能になります。



信頼性の向上

ガバナンス・透明性・コンプライアンスを強化し、データの信頼性を高めます。これにより、確信を持ってデータ主導型の意思決定ができるようになります。データの利用範囲や種類の所有権を明確にします。これにより、データと分析の関係者同士の連携や説明責任を促進し、企業のデータガバナンスの体制を強化することができます。

データ製品の概要

連携された方法は、データ製品の作成を始めるのに最適な方法です。多くの組織が、制御からコラボレーションへ、プロジェクトから製品を生む方向へ移行し、俊敏な組織に生まれ変わることで、データから大きな価値の創出が可能になります。[データメッシュ](#)と[データファブリック](#)は、異なる方法でデータを管理します。双方を組み合わせることで強力な相乗効果を生み出し、データ製品の潜在能力を最大限に引き出すことができます。データメッシュの概念では、データの所有権を分散し、データを使用するチームが所有・管理します。データメッシュの概念において、データ製品は最も明確な効果を発揮する要素です。一方、データファブリックは、最適な方法でビジネス部門にデータ製品を提供する基盤となるデータ管理アーキテクチャです。²



Gartner 社の「CDAO Agenda Survey」では、回答した企業の 50% が既にデータ製品を導入しており、29% が今後 1 年以内に試験導入または本格導入を検討していることが明らかになっています。³

データ製品とは何か?データ製品は、即座に使用できるよう整理された、AI 対応の質の高いデータセットを提供します。ユーザーは、データセットに容易にアクセスし、全社のあらゆるビジネスユースケースに適用することができます。データ製品は、データ用のマイクロサービスとして考えることができます。マイクロサービスは、より小さく状況に基づいた構成ができる単位にサービスを分割して組み合わせることで、アプリケーション機能を提供します。同様に、データ製品は、サービスインターフェースを介して業務領域に応じた関連データへのアクセスを提供します。データ製品は、以下の基本原則に従う必要があります。



業務領域と所有権

すべてのデータ製品は、特定の業務に合わせて調整が必要です。また、データ製品がもたらす価値の提供と展開範囲の拡大に対して責任を負うデータ所有者(個人またはチーム)の任命が必要です。



発見可能性と探索可能性

データ製品カタログにデータ製品を登録し、データ利用者がデータを容易に探索して発見できるようにします。また、データ自体の詳細情報(アクセス方法・ビジネス関連・系統・品質)の提供が必要です。



製品思考

データ製品を作成する際は、所有権・バージョン管理・顧客(ここではデータ利用者)といった製品思考をデータに適用します。データ製品の使用パターンを把握して利用者の意見を組み込むことで、データ製品を強化します。物理的な製品と同様に、データ製品にもライフサイクルがあります。



再利用性

データ製品は、一度構築すると複数のユースケースで再利用することができます。既存のデータ製品から新しいデータ製品を構築することも可能です。



標準化と相互運用性

データ製品は、標準の形式・スキーマ・メタデータの規則に準拠している必要があります。これにより、データエコシステムの他のコンポーネントとの相互運用性と容易な統合を促進することができます。

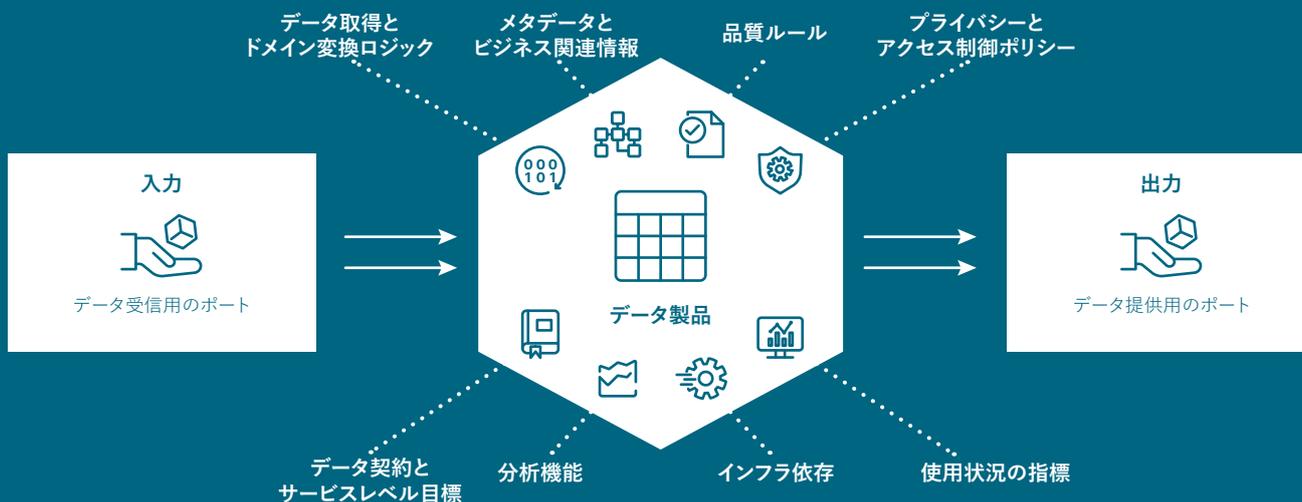


アクセシビリティ

データ製品は、作成時に定義・公開される SQL や REST API などの標準のインターフェースセットを介してアクセスできます。こうしたインターフェースは、データ製品の複雑性を軽減します。

² Gartner 社「6 Lessons Data Leaders Can Learn from Adopters of Data Mesh」, Robert Thanaraj, et al. 2024年 3月 4日 GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

³ Gartner 社「Gartner Chief Data and Analytics Officer Agenda Survey for 2024」2024年 3月 12日 本調査結果は、世界的な調査結果や市場全体を代表するものではなく、調査対象の回答者や企業の感情を反映したものです。GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.



データ製品は、自己完結型の量子です。データ・変換・コード・ガバナンスポリシー・SLA・アクセスパターン・インフラの依存関係を、1単位としてカプセル化したものです。上図は、データ製品の構造を示しています。

- 1 データセット:** データ製品の中核です。テーブル・ファイル・グラフ・ベクトル表現などがあります。データ製品内には、複数の層が存在する可能性があります。ソースシステムから取り込まれた生のデータセット、クレンジング・統合・整理されたデータセット、使いやすさを考慮してモデル化された即座に使用できるデータセットです。
- 2 入力:** データの入力ポートまたは入力ソースです。データの取り込みが必要な SAP アプリケーション内のテーブル、ソースとして使用する必要がある Snowflake 内のテーブル、新しいデータ製品の派生元となる別のデータ製品、リアルタイムで流れるデータを溜めておくメッセージキューなどがあります。
- 3 出力:** データ利用者にデータを提供します。データ製品は、SQL / REST API / ベクトル検索 / メッセージ / ファイルなど、さまざまな標準のインターフェースを公開することができます。また、出力ポートは、データ製品が提供するサービスの種類によって異なります。過去のデータへのアクセスを提供する場合もあれば、データの最新の簡易情報へのアクセスを提供する場合もあります。
- 4 データ取得とドメイン変換ロジック:** データを取得するためのデータパイプラインのコードとロジックです。ソースデータベースからクラウドデータプラットフォームへのデータの移行、テーブルセットの読み込みなどがあります。取得したデータに変換ロジックを適用し、ソース形式から対象のデータセット形式(テーブル・表示・ベクトル埋め込みなど)に変換します。
- 5 メタデータとビジネス関連情報:** 技術メタデータ(テーブル名・ファイル名・カラム定義)と、ビジネスメタデータ(業務領域・製品名・製品の種類・目的・所有権・タグなど)を組み合わせ、データ製品の詳細を把握できるようにします。
- 6 データ品質ルール:** データ製品に対して必要なさまざまなデータ品質の評価を定義します。各ルールには、実行日・ルールの適用範囲・一括または段階的な実行といった設定情報があります。データ品質ルールを規定のスケジュールで実行することで、データの正確性と信頼性を確保することができます。
- 7 プライバシーとアクセス制御ポリシー:** データ製品に、データセットへのアクセスを制御するポリシーを確立します。データセットに個人情報が含まれている場合などに適用します。プライバシーポリシーを確立することで、該当する情報を非表示にしたり、特定のユーザーグループのみにアクセスを許可することができます。
- 8 データ契約とサービスレベル目標:** データ製品で実現すべきサービスの目標値を定義します。たとえば、データ品質の精度指標が 95% 以上、データの鮮度が 5 分置きという要求を定義します。
- 9 分析関数:** 追加のカスタムビジネスロジックをカプセル化して、関数を再利用できるようにします。たとえば、延滞注文という販売注文に関するデータ製品の関数を使用して、複雑なドメインロジックをカプセル化し、延滞している注文を把握することができます。分析関数は、API などの出力ポートを介して公開可能です。
- 10 インフラ依存:** データ製品と紐づいたデータセットのインフラ要件に関する情報です。データセットは、Snowflake や Databricks などのデータプラットフォームに保存されています。変換ロジックなどの関連コードを実行するには、SQL などのコンピュート環境が必要です。

データ製品のライフサイクル

データ製品は、一度作成すればそれで終わりではありません。物理的な製品と同様に、継続的な管理と監視が必要です。各部門のデータチームが、各データ製品の所有と管理を担います。データチームは、価値の提供・導入の拡大・ライフサイクルの維持に対して責任を負います。



以下は、データ製品のライフサイクルにおける主な段階です。

- 1 概念化:** ユーザーやユースケース、ユーザーによるデータの一般的な使用方法を理解するといった、業務領域に応じたビジネスニーズを特定します。
- 2 設計:** ソリューションの仕様を決定します。必要なデータセットとソースの特定、データ製品の特性(機能・双方向性・提供方法など)を定義します。
- 3 精製:** 設計の完了後、データ製品の技術的なインフラと機能を構築します。具体的には、ソースから生のデータセットを収集して統合し、データ品質を確保して変換します。
- 4 展開:** データ製品にアクセスできる状態にします。製品の文書化、所有権の割り振り、既存のビジネスシステムや利用手順との統合、データ製品を有効化してシームレスにアクセスできるようにします。
- 5 公開:** データ製品を有効化した後は、製品の認知度を高めて、利用を促進することが重要です。利用者向けのトレーニングを実施したり、利用者の意見を収集するコミュニケーションの場を構築します。
- 6 使用:** データ製品を市場で販売すると、利用者は、データ製品を探索・発見・理解・アクセスできるようになります。ここでは、使用状況の監視と利用者の意見の収集に重点を置きます。収集する利用者の意見とは、データ製品の用途やパフォーマンス、業務領域のビジネスニーズに合致しているかといった情報が一般的です。
- 7 保守:** データ製品には更新が必要です。継続的に価値を提供するには、改善していく必要があります。収集した利用者の意見とビジネスニーズの変化に基づいて、バージョンアップの際に追加する新たな特性や機能、機能性を提示します。

ライフサイクルの段階に沿って、実用的なインサイトの獲得・データ主導型の意思決定の促進・ビジネス目標の達成を可能にするデータ製品を作成することができます。

データ製品と類似用語の違い

数多くのデータに関する用語がありますが、使用方法が一貫しているとは限りません。「データ製品」という用語は、「データセット」や「データカタログ」と混同されがちですが、同義ではありません。ここでは、こうした概念の基本的な違いについて解説します。



データセットは、多くのデータ製品を構成する基本的な要素の1つです。一方、データ製品は、生のデータセットだけでなく、さまざまな要素を含んでいます。

データセット：原材料

一般的には、整理された一連の構造化・非構造化データポイントを指します。生データ（顧客の詳細情報・財務記録・センサーからの読み取り値など）が含まれる場合があります。データセットを木材などの原材料だとすると、原材料を有効活用するには、まず精製が必要です。データセットとデータ製品の大きな違いの1つは、データセットには業務領域に特化したビジネス情報がありません。

データ製品：完成品

多くのデータ製品は入念に整理・処理されたデータセットを使用し、前述のデータセットによる成果物と組み合わせて価値を提供します。データセットが木材なら、データ製品は木材で組み立てた家具のようなものです。

以下は、データセットとデータ製品の違いです。

- **ビジネス状況業務領域を重視：**データ製品は、特定のビジネスニーズやユーザーの要求に応じてカスタマイズできます。生データの情報だけでなく、定義・系統・品質評価などの情報も含まれます。
- **利便性とアクセシビリティ：**データ製品は、価値や利便性を向上するよう設計されています。技術面に精通していないビジネス部門のデータ利用者が容易に使用できるよう、整理・パッケージ化されています。
- **実用的なインサイト：**データ製品は、単にデータを提示するだけでなく、運用・分析・予測・規範に関するインサイトも提供します。可視化されたダッシュボード・構築済みの分析モデル・SQL / API を介した他のツールとの統合から、インサイトを取得することができます。
- **所有権と説明責任：**データ製品には明確な所有権があります。これにより、社内における開発や使用、意見提示のあらゆる側面に対する説明責任を確保します。データ製品の所有権は、ビジネスの状況によって異なります。データ所有者は、データ製品の作成者、関連するユースケースを担当するビジネス部門の利用者、新たな職務であるデータ製品マネージャーの場合があります。



データ製品は、データセットを、情報に基づいた意思決定とビジネスの成長を促進する使用可能な価値の高い資産に変換します。Harvard Business Review は、「製品としてデータを扱う企業は、新たなユースケースにデータを実装する時間を最大 90% 短縮できる。また、総保有コスト（技術・開発・保守）を最大 30% 削減し、リスクやデータガバナンスの負担を軽減できる」と予測しています。⁴

⁴[A Better Way to Put Your Data to Work] Harvard Business Review 2022年7月 - 8月

データ製品カタログ：完成品の一元管理

データ製品カタログは、データ製品と関連メタデータ（詳細・タグ・バージョン・所有権ラベルなど）を格納する包括的なリポジトリとして機能します。これにより、データ製品の検出と理解をサポートします。また、データの作成者と利用者のニーズに合致するカスタマイズされたソリューションを提供します。データ製品マネージャー向けには、データ製品のライフサイクル全体（作成から有効化、使用終了まで）を構成・有効化・管理できるツールを提供します。利用者向けには、オンラインショッピングのようにデータ製品を利用できる場を提供します。

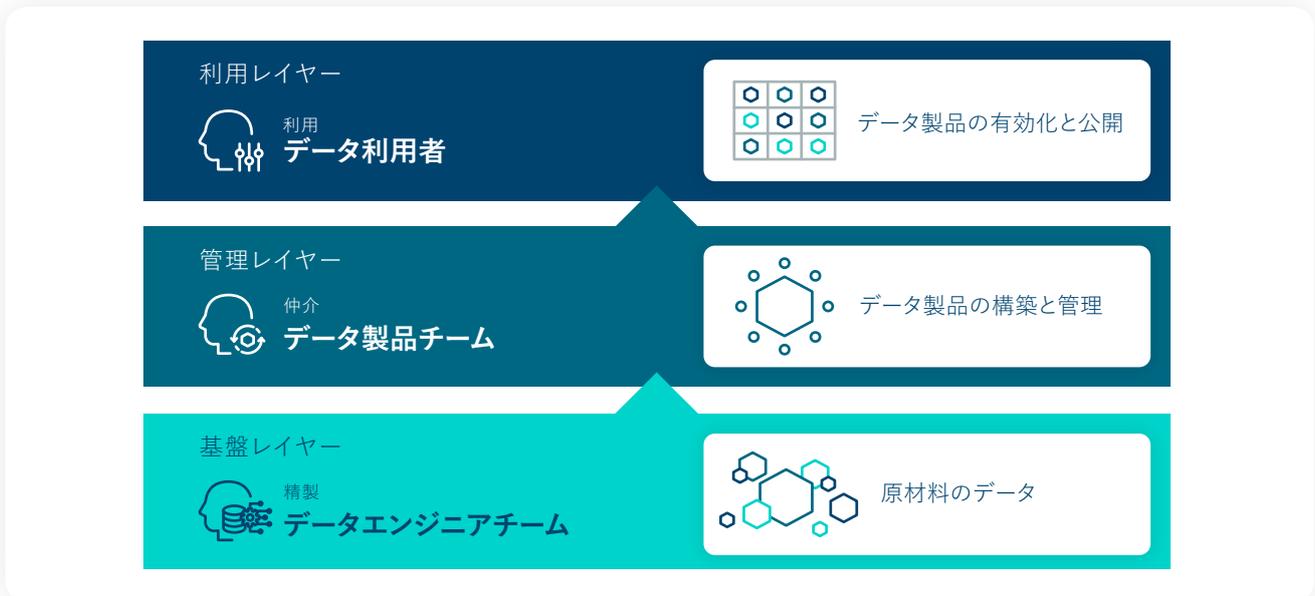
以下は、データ製品カタログとデータカタログの違いです。

項目	データカタログ	データ製品カタログ
役割	データガバナンスの強化	データ価値の向上
範囲	全社向け	業務領域向け
管理モデル	一元管理	連携管理
目的	探索可能性・コンプライアンス・透明性	利便性・信頼性・実行可能性
内容	多様なメタデータ (すべてのデータ資産に関する技術・ 運用・ビジネス情報など)	信頼できるデータ製品・データ製品の ライフサイクルに関するメタデータ・ データ製品の記述と所有権
単位	数十万単位	数百単位
対象	主にデータガバナンス	データ作成者・利用者

データ管理における課題は、データプラットフォームの構築と効果的な導入です。ほとんどの企業は、数億円もの多額の投資をしてデータインフラを構築しますが、**データの作成者と利用者**の間に生じる**データ共有の障壁**という深刻な課題に直面することになります。その理由は、多くのテクノロジープラットフォームが生成する結果が信頼性に欠けており、データ利用者のビジネス目標に合致していないからです。**解決するには、データカタログとデータ製品カタログの両方の導入が必要です。**スーパーマーケットの売り場に例えると、データカタログは、小麦粉・バター・砂糖などの原材料です。一方、データ製品カタログは、ケーキやパンなどの完成品になります。

データ製品カタログ:レイヤーと関係者

データ製品カタログは、自社のデータ製品を整理するための主な基盤です。3つのレイヤーに分かれており、各レイヤーは異なる関係者に合わせて調整されています。データエンジニア・データスチュワード・データアナリスト・新たな職務(データ製品マネージャーなど)を問わず、各グループはそれぞれの目的と要件に応じて、独自の方法で運用します。この方法は、ユーザーのニーズに基づいて課題を分けて適切なレベルを抽象化することで、容易な導入と利用を可能にします。



基盤レイヤー

データ製品カタログの基盤は、技術インフラを確立し、後続のレイヤーをサポートする機能を提供する重要なレイヤーです。

以下は、基盤レイヤーの主な機能です。

- **データの精製:** ファイル・クラウドデータウェアハウス・データレイク・リアルタイムデータストリームなど、あらゆるソースからデータを移動して処理します。
- **セキュリティとアクセス制御:** データセキュリティ制御を適用してデータを保護し、許可されたユーザーのみにデータへのアクセス権を付与します。
- **メタデータ管理:** データ系統・データ辞書・スキーマなど、データに関する有益な情報を収集して整理します。
- **データ品質と系統の追跡:** データの品質を監視して正確性を確保し、データの系統を追跡して信頼性を確保します。

基盤レイヤーは、データエンジニア、データスチュワード、データアーキテクトなどの主な技術職が所属するデータプラットフォームチームをサポートします。このレイヤーの役割は、基盤レイヤーの機能をセルフサービスとして構築し、技術者以外のユーザーが容易に起動できるようにすることです。基盤レイヤーに必要な機能は、メタデータ管理・データの取り込み・データの変換・データ品質・データセキュリティ向けのさまざまなツールに搭載されています。

管理レイヤー

管理レイヤーは、新たな概念の1つです。新たな職務のデータ製品マネージャーまたはチームのニーズに応える必要があります。業務領域に適した信頼できるデータ製品を開発・管理・保守ができるようサポートします。データ資産の一覧作成・移行・変換・保護は、基盤レイヤーのサービスに依存します。データ利用者が、基盤レイヤーに格納されている生のメタデータから実用的なインサイトを獲得できるよう、橋渡しの役割を担います。

以下は、管理レイヤーの主な機能です。

- **データ製品の構成:** データ製品の設計、入力と出力の定義、あらゆる成果物を構成する機能を提供します。
- **データ製品の展開:** データのバージョンを管理し、データが使用できる状態になった時に、利用者向けにデータ製品を有効化します。
- **データ製品の保守:** 使用状況の追跡、利用者の意見を収集、変更や更新、データ製品のライフサイクルにおける各段階を把握します。

データ製品の管理レイヤーは、主に製品管理チーム、特にデータ製品のライフサイクルを管理するデータ製品マネージャーのビジネス目標の達成をサポートします。

利用レイヤー

利用レイヤーは、データ製品カタログの最上位に位置し、システムにおけるデータ利用者向けの層です。セルフサービスのデータ探索と利用を可能にし、本社におけるデータ主導型文化を促進します。

以下は、利用レイヤーの主な機能です。

- **探索と発見:** 直観的なインターフェースで、ユーザー自身のニーズに最適なデータ製品を容易に探索・発見することができます。
- **学習と信頼性:** データ利用者は、メタデータとサンプルデータの両方からデータ製品を理解し、サンプルコードとデータの記述を参照して利用法を学習することができます。また、データの品質やシステムを確認することで、安心してデータ製品を使用することができます。
- **アクセスと権限:** 特定のデータ製品へのアクセス権を限定し、データセキュリティとデータガバナンスのポリシーの遵守を確保します。
- **データ利用インターフェース:** API や可視化ツール、データベースへの直接的なアクセスポイントを介して、ユーザーがデータ製品を使用できるようにします。一部の製品では、ユーザーの多様なニーズやスキルセットに対応できるよう、複数のインターフェースの利用が可能な場合があります。

利用レイヤーは、主にビジネスアナリスト、データサイエンティスト、基幹業務ユーザー、その他の主な意思決定者をサポートします。

データエコシステムのすべての関係者をサポートするデータ製品のメリット

データは企業の重要な資産です。データの真の可能性を解放するには、データへの容易なアクセス、データへの深い理解と効果的な使用が不可欠です。データ製品は、この資産を全社で利用できるよう、生データを利用可能な整理された状態に変換し、この課題を解決します。

ここでは、データ製品がデータエコシステムの主な関係者に与える影響を解説します。



データエンジニアリングチームにとって、データ製品は明確に定義された共通言語です。データ製品は、データの作成者と利用者間のコミュニケーションを構築し、明確にします。これにより、データ作成者は再利用可能なデータ製品を何度も開発できるようになり、総保有コストを削減することができます。さらに、一元管理された大規模かつ複雑なデータパイプラインを構築する必要がなくなり、各製品を実装できる小規模かつ独立したパイプラインの開発に専念できるようになります。



データガバナンスチームにとって、データ製品は中核の監査ポイントです。製品の信頼性を確保し、適切に統制することで、容易にリスクを最小化することができます。ポリシーと標準を一元的に定義し、すべてのデータ製品チームが連携された方法でポリシーを適用します。



データ製品マネージャーにとって、データ製品はデータ作成者と利用者の連携を強化するものです。また、データエコシステム全体でコラボレーションと説明責任を促進し、データの所有権と説明責任を確保することができます。



データ利用者にとって、データ製品は業務領域に適したデータ資産です。必要な情報を確実に提供し、ビジネス価値の最大化をサポートしてくれます。また、作業のやり直しを低減し、より迅速に成果を挙げるすることができます。

関係者同士が協力してデータ製品カタログを効率的に活用し、データ資産の価値を最大化することができます。これにより、データ主導型の意思決定を促進し、ビジネス目標の達成を可能にします。

住宅ローン業界では、データ製品は単なる数値やアルゴリズムではなく、変革の原動力です。整理されたセルフサービス型のデータ製品を定義することで、融資担当者が適切な借り手に最適な住宅ローン計画を提案できるようにしたいです。また、財務部門がよりの確なコスト評価とリスク管理ができるようにしたいと考えています。当社の事業でデータ製品を活用することは、住宅ローンの未来を真に再定義することになるでしょう。

Julia Fryk 氏

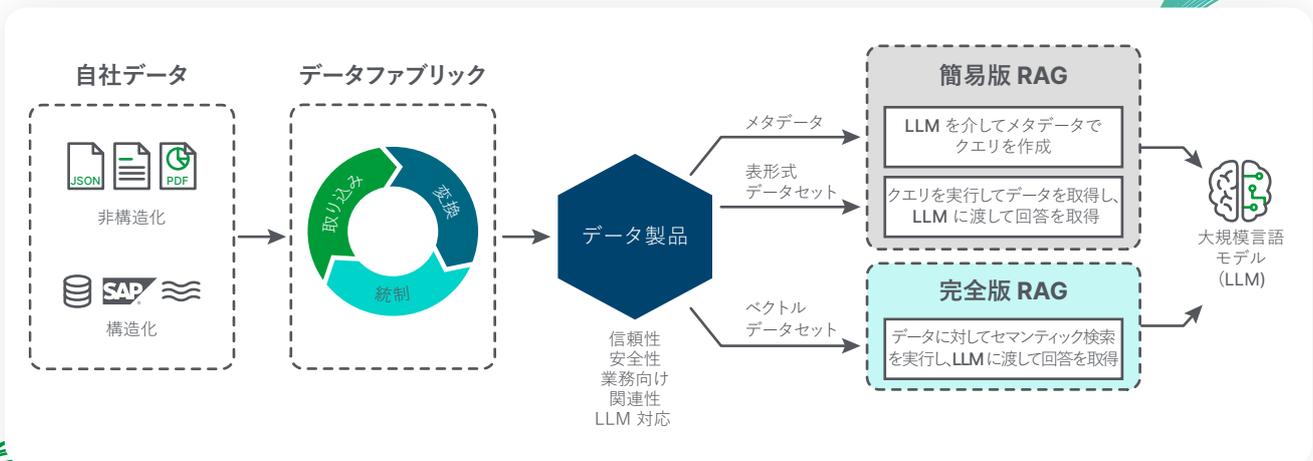
Waterstone Mortgage 社 データアーキテクト最高責任者

データ製品で大規模言語モデル(LLM)にコンテキストデータを提供

生成 AI は、あらゆるビジネスの生産性を高める大きなチャンスを生み出しています。生成 AI の中核となるのは、Web ページや文書といった膨大なテキストデータコーパスで事前学習された、LLM(GPT-3 / Anthropic / Llama / BERT)などの基盤モデルです。基盤となる LLM は自社固有のデータを把握していないため、自社データで LLM に学習させる必要があります。

自社データを LLM に提供する最も効果的な方法の 1 つは、検索拡張生成(RAG)と呼ばれるテクノロジーを使用することです。まず、ユーザープロンプト(質問)に関連するコンテキストデータを**取得**します。このデータでユーザープロンプトを**補強**し、LLM に渡して回答を生成します。RAG には、簡易版と完全版の 2 種類があります。

- 簡易版 RAG(構造化データ向け)では、RAG アプリケーションがメタデータを LLM に渡し、SQL クエリを生成します。次に、表形式のデータセットに対してクエリを実行し、その結果を LLM に渡して回答を生成します。この手法は、データから明示的な回答を取得する必要がある場合に使用されます。
- 完全版 RAG(構造化・非構造化データ向け)では、まずデータをベクトル化します。次に、RAG アプリケーションは、ベクトル化されたデータに対してセマンティック検索を実行し、データを取得します。取得したデータを LLM に渡して回答を生成します。



どちらの方法も、RAG で利用可能な高品質かつクリーンなコンテキストデータが必要です。**データ製品は、あらゆる方法でコンテキストデータと業務領域に適したデータを提供します。**データ製品には、豊富なメタデータ(ビジネス・技術・業務向け)と、利用者向けに高度に整理されたデータセット(ファイル・表・ベクトル化)があります。そのため、データ製品は、RAG ベースのアプリケーションを構築する最適な手段となります。

RAG アプリケーションの構築は、まずデータ製品カタログやマーケットプレイス上で、提示された質問に基づいた適切なデータ製品を探すことから始まります。売上に関する質問の場合、データ製品と紐づいた詳細なメタデータに基づいた CRM の「案件」を示すデータ製品が最適かもしれません。注文に関する質問の場合、データ製品と紐づいた詳細なメタデータに基づいた SAP の「販売注文」を示すデータ製品が最適かもしれません。データ製品を特定したら、RAG アプリケーションはユースケースに基づいて、データ製品のデータセット(表またはベクトル化)に直接クエリすることができます。

まとめ

データは、今日のビジネスを牽引し続けています。データ製品は、データの信頼性とガバナンスを強化し、データの価値を最大化してビジネス成果を加速する上で、ますます重要になっています。Qlik は、セルフサービスのデータ基盤の構築を全力でサポートします。Qlik は、データチームがデータ製品を構築・管理し、データ利用者がデータ製品を迅速に探索・発見・活用できるようにします。最終的に、企業はビジネス成果を促進することができます。

データ製品は強力なツールですが、あらゆる課題を解決する万能なソリューションではありません。データ製品の真のパワーを活用するには、ビジネスニーズの特定、明確に定義した連携データ戦略の策定、適切なデータガバナンスの確保、チーム間のコラボレーションの促進が不可欠です。

データの統合や品質だけでなく、業務領域に適したデータ製品の開発が求められています。Qlik のデータ品質およびガバナンスは、データ製品の概念の具現化に重点を置いています。今日の顧客のニーズに寄り添う姿勢は、同社のサービスが十分に成熟していると言えるでしょう。

Stewart Bond 氏

IDC 社 データインテリジェンス・統合ソフトウェア部門長

自社データの価値を
最大化するには？

詳細を見る



Qlik について

Qlik は、複雑なデータ状況を実用的なインサイトに変換し、戦略的なビジネス成果を促進します。世界 40,000 社以上の顧客にサービスを提供している Qlik の製品ポートフォリオは、最先端かつエンタープライズ水準の AI / 機械学習と広範なデータ品質を基盤としています。また、優れたデータ統合およびデータ統制、多様なデータソースに対応する包括的なソリューションを提供します。Qlik の直感的でリアルタイムの分析は、隠れたパターンの発見や複雑なビジネス課題の解決、新たなビジネスチャンスの獲得を支援します。さらに、実用的で高度な拡張性を備えた Qlik の AI / 機械学習ツールで、適切で迅速な意思決定を可能にします。Qlik は顧客の戦略的パートナーとして、プラットフォームに依存しないテクノロジーと専門知識で、顧客の競争力を高めます。

qlik.com